International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Nonparametric Statistics: A Mixtures of Finite Polya Trees

Asri Ode Samura

Tadris Matematika, IAIN Ternate

Abstract: ***This Writing be explained an approach in nonparametric Bayesian, a mixture of Polya tree (MPT). MPT uses a partition of supports from a density of its original distribution. In general, the density keeps the original form of the distribution in every mixture of partition as well as adds a new parameter obtained from conditional probabilities. A model of Polya tree can be applied widely and also can be programmed easily by giving MCMC scheme to fit the original parametric model.***

Keywords: Bayesian; A generalized liniear mixed model; GLMM

1. Introduction

Distribution sampling parameters is an analyzing which needed a prior probability by Bayes analyzed, it would be hard to choose a logically in a lot of parameters complications, than it would be divided in two groups appropriate: (1) it has a correlate to relatively original parametric, and (2) it has a correlate to its relatively original generation. Standard method could to use to a prior shaping to an original parameters,as an approach in a model of linear generalization; see Bedrick, Christensen and Johnson (1996). An exactly prior indicates existing in parameters from a generalization method.Polya tree diagram is a random probability distribution. An original parametric component which has decided by a condition (such as; a normal distribution within approximately μand capriceσ^2), the generalization of distribution component with the prior indicator is unlimited Polya tree diagram. The integratedpriorto the original component parameter (sample; μ dan σ^2) are shaping to composite Polya tree diagram.

MPT is used to a universal method, but in example we would entangling tousle the mixing model of linear generalization (GLMM), and the both of it would be involve a normal distribution generalization which is collective assumption to its, see Breslew and Clayton (1993). The notoriously framework to longitude measurement analyze and appearing data of the groups in many field like agriculture, biology, epidemiology, economic and geo-physics. Those models are numerously a correlation among in a group of observations by randomize effect input to the linear prediction component from these models. Even especially the GLMM adjustment characteristic is complexity, randomize intercepts standard and intercept models/randomize slot which has randomize effect to presently a normal distribution could be compatibility routine into commercial packets software such as; SAS and Stata. These models are flexibility into behavioral heterogenic accommodation, but meanwhile they have the same shortage which has endure decrease to an assumption purpose from other statistic models of Gauss distribution.

As a wellness strategic taken into unsuitable normalize assumption into more flexibility distribution assumption to the models randomize effect. Than similarly non parametric built up from GLMM parametric.

A kept strategic to normalize assumption is unsuitable which flexibility to distribution assumption input to models of randomize effect. Even if the non-parametric built up from GLMM is needed. We could explain these wide within unusual effect to the model traditional assumption into GLMM by an example our reality life. As a comparison purpose, we have to adapting to all of models suitable with the both of randomize normal distribution and generalization.

2. Base Theoretics

2.1 Nonparametric Statistic

A procedure of statistically is a part from a type of nonparametric even if has an indefinite characteristics which is a logical improvement when some of assumptions has a general characteristic, such as using the generally characteristic is a function to researching for non-parametric characteristic from a statistically procedure working by Walsh (1962).

Limiting ‘parametric and non-parametric adjectives to the statistically hypothesis should be useful to statically, test, or infers type. These ways are probably smart, but it had makes many confusing are arrive from useful statement like ‘non-parametric test’ and infers non-parametric. By Kendall, and Sundrum (1953)

Definitions 2.1.1: *As statically method is non-parametric if there are not requiring to ones of criteria below:*

- A statistically method is used to the data measurement nominal scales.
- These methods are used to the data measurement nominal ordinals.
- These methods are used to the data measurement interval or ratio, where this function is randomize variable distribution wouldn’t be resultespecial data except unlimited numerous from the unknown parameter.

2.2 Probabilistic Theory

Depend on extend a sample space Ω from a randomize experiment and a σ-fields$\varepsilon \subseteq \Omega$.When ε in compilation is called phenomenon. In $A \in \varepsilon$phenomenon setting up by a real occupation number P (A) which has a size numerously from a resource experiment result and it would be as component of A. An event probability could be interpretation suitable within main concept of relative frequency and it could be observation by an axiomatic method.

Definition 2.2.1 *a function of P is definitely use to the real occupation number component which is called probability measurement if P require to: (i)* $0 \leq P(A) \leq 1$ *for every* $A \in \varepsilon$. *(ii)* $P(\Omega) = 1$. *(iii) jika* $A_1, A_2, A_3, \dots$ *...Is liberating phenomenon into* ε, *so formulated to its' would be:*

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i), \tag{2.1}$$

The real occupation number of P (A) is called as phenomenon of probability (Degroot, 1970, Rohatgi.1976).

Definition 2.2.2: *for example A and B are phenomenon in* ε *with P(A) > 0. When the probability of B if given by A within notation P(B|A), it formulated would be:*

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \tag{2.2}$$

Theorems 2.2.1: *For every phenomenon A and B, so it would be:*

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

An independence from A to B if B phenomenon would not give information to Aor A and B as the phenomenon difference, so it would be:

- $P(A|B) = P(A)$or
- $P(B|A) = P(B)$or also
- $P(A \cap B) = P(A)P(B)$.

Theorems 2.2.2: *If A and B are called independence so it would be:*

- $A^c$and B independence.
- AandB^cindependence.
- A^candB^cindependence.

2.3 Conditions Of Randomize Variable Probability

If X is randomize variable, so "X =x" is suitable with some of phenomenon into sample space,it means that phenomenon resource from point to all of occupation number componentis given to value "x." Than"X =x" sometimes called as "X=x,phenomenon" it mean the effect of phenomenon settled from xvalue is source from randomize Xvariable."

Definition 2.3.1: *Probability condition from Xwhich given byY, symbolized as P(X =x|Y=y), is randomize variable probability Xisx value, and it also given by Y randomize variable as suspicion y. it symbolized would be:*

$$P(X = x \mid Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)} \quad \text{if} \quad P(Y=y) > 0 \tag{2.3}$$

Function of probability to randomize variable *X*, symbolize used is *f(x)*, it mean the *X* probability function would giving assumption to *x* value, for every real occupation number, another symbolize is,

$$f(x) = P(X = x) \tag{2.4}$$

Function of Probability is 0 to the *x*value so*X*could not assumptive.

Definition 2.3.2: *Distribution function from one of randomize variable of X, which symbolize used is f(x) is function of X probability less than or similarly to every real occupation number of x, below its symbolized:*

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \tag{2.5}$$

For example randomize variable of X, binomialdistribution is probability distribution.Its symbolized is,

$$f(x) = P(X = x) = \binom{n}{x} P^x q^{n-x} x = 0,1,2,\dots,n \tag{2.6}$$

Where n is the positive spherical occupations number, $0 \leq P \leq 1$and$q = 1 - P$. And the symbol of its used is $0! = 1$. So, sustainability distribution function is

$$F(x) = P(X \leq x) = \sum_{i \leq x} \binom{n}{1} P^i q^{n-i} \tag{2.7}$$

2.4 Characteristics of Randomize Variable

The characteristic randomize variable is similarly like as the function of probability and function of distribution. Probability function is shown to all of every properties from randomize variable, caused probability function would show the randomize variable from phenomenon approximately. And has correlate to every probability values.

Generally methods are used given to randomize distribution variable, it means that it would give some of select distance from randomize variable. They are called quintile, median, quartile, deciliter and percentile are useful into randomize variable.

Definition 2.4.1: *Give to randomsize variable of Xto the probability function off(x), and take an example u(X) is the rill function value from X.than the percentage prize of u(X), could be symbolize as*$E[u(X)]$*or given to symbolize like:*

$$E[u(X)] = \begin{cases} \sum_x u(x)f(x) & x = \mathit{diskrit}. \\ \int u(x)f(x)\,dx, & x = \mathit{kontinu}. \end{cases} \tag{2.8}$$

Ordinarily meanX, symbolized asμ, than

$$\mu = E(X) \tag{2.9}$$

Even if could be symbolize write on is

$$\mu = E(X) = \sum_x x f(x) \tag{2.10}$$

Definition 2.4.2: for example randomize variable *Xused by mean* μ*and probability function of f(x). XVarians, which is symbolize to* σ^2*orVar(X),is:*

$$\sigma^2 = E[(X - \mu)^2] \tag{2.11}$$

Variance of X also could be symbolize as

$$\begin{aligned} \sigma^2 &= \sum_x (x-\mu)^2 f(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2) f(x) \\ &= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x) \end{aligned}$$

Where$\sum_x f(x) = 1$, and using symbolize to(2.17), so the equality is:

$$\sigma^2 = E(x^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2 \tag{2.12}$$

Definition2.4.3: For examples $X_1, X_2, \dots, X_n$*Randomize variable within probability function to*$f(x_1, x_2, \dots, x_n)$, *than use the examples tou*$(X_1, X_2, \dots, X_n)$*are riilfrom* $X_1, X_2, \dots, X_n$. *So, the percentage prizes fromu*$(X_1, X_2, \dots, X_n)$*is:*

$$E[u(X_1, X_2, \dots, X_n)] = \sum u(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) \quad (2.13)$$

2.5 Principles of Bayesian Inferential

2.5.1 Bayes Literatures

Bayes has special characteristics which has suitable indicators to scientific research. Most of these points are:

1) Bayes analyze use to automatically resourcing from data information's entire.
2) Unaccepting inferential should resource from unsuitable assumption and it is not enough from inferential system.Therefore part of most models includes first step distribution suitable would be used to the models explain.
3) This inferential system is ready to use into most of opportunity models. So, high concentrating to use this models are short into models of mathematic method, and more than it, most should has to build up to useful of scientifically.
4) Anomaly found into researching of theoretically sampling, estimate to selecting concentration and confidential interval.
5) Bayes inferential give to suitable ways to explicit improving and kept the assumption to Scientifics and non-first Scientifics.

2.5.2 BayesTheorems

For example y′ = (y_1,…,y_n) is a vector from n observation to opportunity distribution p(y|θ) its depend on to k value parameters θ′ = (θ_1, …, θ_k). And for examples that θ has an opportunity distribution p(θ). So, it's symbolize is:

$$P(y|\theta)P(\theta) = P(y,\theta) = P(\theta|y)P(y) \quad (2.14)$$

Observation data given by y, condition of distribution from θwould be:

$$P(\theta|y) = \frac{P(y|\theta)P(y)}{P(\theta)}. \quad (2.15)$$

Than

$$P(y) = EP(y|\theta) = C^{-1} = \begin{cases} \int P(y|\theta)P(\theta)d\theta\ \theta kontinu \\ \sum P(y|\theta)P(\theta)\theta discrit \end{cases} \quad (2.16)$$

Where sigma and its integrals are taking to estimate from θ and it would accepting, and then could be accepting E[f(θ)] for mathematic theory expectations from f(θ) has correlate to distribution P(θ). Therefore these symbolize are below:

$$P(\theta|y) = cP(y|\theta)P(\theta). \quad (2.17)$$

2.5.3 PosteriorDistributions

Posterior distributionθgive to$x$(or it's called posterior) would be symbolize as $\pi(\theta|x)$, And its definition as a condition of distribution fromθ,than give to $x$as observation sample.Or the other way is $\theta$and$X$has a density or subjective.

$$h(x|\theta) = \pi(\theta) f(x|\theta), \quad (2.18)$$

And X has a marginally (without right condition)and its density is:

$$m(x) = \begin{cases} \int f(x|\theta)\pi(\theta)d\theta \quad jika\ \theta\ kontinu \\ \sum_\theta f(x|\theta)\pi(\theta) \quad jika\ \theta\ diskrit \end{cases} \quad (2.19)$$

It's method already clearly that (giving to $m(x) \neq 0$)

$$\pi(\theta|x) = \frac{h(x|\theta)}{m(x)}. \quad (2.20)$$

Finding out for estimator bayes, so the useful method is

$$E(\pi(\theta|x)) = \begin{cases} \int \theta f(x|\theta)\pi(\theta)d\theta jika\theta kontinu \\ \sum_\theta \theta f(x|\theta)\pi(\theta) jika\theta diskrit \end{cases} \quad (2.21)$$

Estimating to Bayes estimator use the method of Box-Tiao.From the equality of (2.14) and this symbolize is write on to below:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \quad (2.22)$$

The constant comparison has selected and has a form of symbolize to $\pi(\theta|x)$be a density.

Theorems 2.5.1: *if$x_1, x_2, \dots, x_n$is a randomize sample from distribution $N(\mu, \lambda^{-1})$within precisions $\lambda, (\lambda > 0)$has known, for example prior distribution for μ averages is the distribution of$N(\mu_0, (n_0\lambda^{-1}))$, than the condition of posterior distribution μmeanwhile giving to $X_i = x_i$ values, for$i = 1,2, \dots, n$is distribution of$N(\mu_n, \lambda_n^{-1})$where is:*

$$\lambda_n = (n_0 + n)\lambda \text{dan} \mu_n = \frac{n\bar{x} + n_0\mu_0}{n_0 + n} \quad (2.23)$$

When it given to $X_i = x_i$, for$i = 1,2, \dots, n$caused of$\lambda$is known, but use the substitutions functions of $L(\mu|\underline{x}) \propto Exp\{-n\lambda(\mu - \bar{x})^2/2\}$and function of prior density$\mu$to$\mu(\mu|\lambda) \propto Exp\{-n_0\lambda(\mu - \mu_0)^2/2\}$into below equality method.

$$\pi(\mu|\underline{x}) \propto L(\theta|\underline{x})\pi(\theta) \quad (2.24)$$

Than gets below method

$$\pi(\mu|\underline{x}) \propto L(\mu|x)\pi(\mu)$$
$$\propto Exp\left\{-\frac{1}{2}[n\lambda(\mu - \bar{x})^2 + n_0\lambda(\mu - \mu_0)^2]\right\} \quad (2.25)$$

Expression $n\lambda(\mu - \bar{x})^2 + n_0\lambda(\mu - \mu_0)^2$as a part of equation right (2.25) would be explained being:

$$n\lambda(\mu - \bar{x})^2 + n_0\lambda(\mu - \mu_0)^2 = (n_0 + n)\lambda(\mu - \mu_n) + \frac{\lambda n n_0(\bar{x} - \mu_0)^2}{n_0 + n} \quad (2.26)$$

Where $\mu_n$is similarly like the equation (2.23), because last quarter of part equation right in (2.26) is constant to μ, so this quarter could be ignored, and after that the substitutions result taking back into equation (2.25) than it would finding out the posterior function of density.

$$\pi(\mu|\underline{x}) \propto Exp\left\{-\left(\frac{1}{2}\right)n\lambda(\mu - \mu_n)^2\right\} \quad (2.27)$$

As a density function from a normally distribution $N(\mu_n, \lambda_n^{-1})$ where$\mu_n$and$\lambda_n$are such as equation in (2.23) even while this theorem is improved.

Theorems: 2.5.2: *if $x_1, x_2, \dots, x_n$Is randomize sample from$N(\mu, \lambda^{-1})$; distribution for $\mu$and$\lambda$and the both of its unknown theorems. If the prior distribution within to (μ, λ)has distribution of $NG(\mu, \lambda; \mu_0, \lambda_0, \alpha_0, \beta_0)$; $-\infty < \mu_0 < \infty, n_0 > 0, \lambda > 0, \alpha > 0, \beta > 0$, than within posterior distribution maka distribusi posterior of$(\mu, \lambda|x)$as distribution of$NG(\mu, \lambda|\mu_n, \lambda_n, \alpha_n, \beta_n)$where these theorems are below:*

$$\mu_n = (n\bar{x} + n_0\mu_0)/(n_0 + n) \quad \text{dan} \quad \lambda_n = (n_0 + n)\lambda \quad (2.28)$$

And

$$\alpha_n = \alpha_0 + {}^{n}/_{2} \ \text{dan} \ \beta_n = \beta_0 + \frac{1}{2}\sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n n_0(\bar{x} - \mu_0)^2}{2(n_0 + n)} \quad (2.29)$$

Improvement:

Given by $X_i = x_i i = 1,2,3, \dots, n$, because $\mu$and$\lambda$the unknown result both of its, and the functions of distributions like hood.

$$L(\mu,\lambda|\underline{x}) \propto \lambda^{n/2} Exp\{-(\lambda/2)\sum_{i=1}^{n}(x_i-\bar{x})^2\} exp\{-n\lambda(\mu-\bar{x})^2/2 \quad (2.30)$$

And the gamma prior functions of density to:

$$\pi(\mu,\lambda|n_0,n_1,\dots,n_n) = \pi(\mu,\lambda|n_0,\mu_0,\alpha_0,\beta_0)$$
$$\propto \left\{\lambda^{1/2} Exp\left\{-\frac{n_0\lambda}{2}(\mu-\mu_0)^2\right\}\right\}(\lambda^{\alpha_0-1}Exp\{-\beta_0\lambda\}) \quad (2.31)$$

Into equality$\pi(\theta|x) \propto L(\theta|x)\pi(\theta)$where $\theta = (\theta_1,\theta_2) = (\mu,\lambda)$, even the resulting is:

$$\pi(\mu,\lambda|\underline{x}) \propto \lambda^{(n/2)+\alpha_0-1}\lambda^{1/2}Exp(-\beta_0\lambda)$$
$$\cdot\, Exp\left\{-\left(\frac{\lambda}{2}\right)\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\mu-\bar{x})^2+n_0(\mu-\mu_0)^2\right]\right\} \quad (2.32)$$

Because $\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\mu-\bar{x})^2+n_0(\mu-$ $\mu 02$explaining could be

$$(n_0+n)(\mu-\mu_n)^2+\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{nn_0(\bar{x}-\mu_0)^2}{n_0+n} \quad (2.33)$$

With $\mu_n$such as to equality (2.28) and equality (2.32) and then could be

$$\pi(\mu,\lambda|\underline{x}) \propto \lambda^{(n/2)+\alpha_0-1}\lambda^{1/2}$$
$$Exp\left\{-\frac{\lambda}{2}(n_0+n)(\mu-\mu_n)^2 + \sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{nn_0(\bar{x}-\mu_0)^2}{n_0+n}\right\}$$
$$Exp(-\beta_0\lambda) \propto \lambda\, Exp\{-\lambda^{1/2}(n_0+n)(\mu-\mu_n)^2/2\}$$
$$\lambda^{(n/2)+\alpha_0-1}Exp\left\{-\left[\beta_0+\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{2}\frac{nn_0(\bar{x}+\mu_0)^2}{n_0+n}\right]\lambda\right\}$$

Or similarly like:

$$\pi(\mu,\lambda|\underline{x}) \propto$$
$$\{\lambda^{1/2}Exp\,[-\lambda_n(\mu-\mu_n)^2/2]\}(\lambda^{\alpha_n-1}Exp(-\beta_n\lambda)) \quad (2.34)$$

In this function is density from normally distribution of gamma $NG(\mu,\lambda;\ \mu_n,\lambda_n,\alpha_n,\beta_n)$with $\mu_n,\lambda_n,\alpha_n,\beta_n$ as explain in equality to (2.28) and (2.29)

Theorems; 2.5.3: *Given to $x_1,x_2,\dots,x_n$from distribution of $N(\mu,\lambda^{-1})$; with$\mu$and$\lambda$and the both if it unknown. If prior informative distribution is used $\pi(\mu,\lambda)=\lambda^{-1}$so it would working with posterior distribution$(\mu,\lambda|\underline{x})$even the distribution is $NG(\mu_n',\lambda_n',\alpha_n',\beta_n')$, where*

$$\mu_n' = \bar{x} \quad \text{dan} \quad \lambda_n' = n\lambda \quad (2.35)$$

And

$$\alpha_n' = (n-1)/2 \quad \text{dan} \quad \beta_n' = \sum_{i=1}^{n}(x_i-\bar{x})^2/2 \quad (2.36)$$

Given to $X_i = x_i$where$i = 1,2,\dots,n$, with$\mu$and$\lambda$the both of these result are unknown$\pi(\theta|\underline{x}) \propto L(\theta|x)\pi(\theta)$could to use for$\underline{\theta} = (\theta_1,\theta_2) = (\mu,\lambda)$into equality $\pi(\theta|\underline{x}) \propto L(\theta|x)\pi(\theta)$this is called like hood distribution function to$L(\mu|\underline{x}) \propto Exp\{-n\lambda(\mu-\bar{x})^2/2\}$, and the functions of density prior uninformative to equality$\pi(\mu,\lambda) = \lambda^{-1}$.So, the result to its is:

$$\pi(\mu,\lambda|\underline{x}) \propto L(\mu,\lambda|\underline{x})\pi(\mu,\lambda)$$
$$\propto \lambda^{-1}\lambda^{n/2}Exp\left\{-\left(\frac{\lambda}{2}\right)\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}\frac{Exp\{-n\lambda(\mu-\lambda)\}^2}{2}$$
$$\propto n^{1/2}\lambda^{1/2}Exp\{-n\lambda(\mu-\bar{x})^2/\}\cdot((n-1)/2)^{-1}$$
$$\cdot\, Exp\left\{-(\lambda/2)\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}$$
$$\pi(\mu,\lambda|\underline{x}) \propto (n\lambda)^{1/2}Exp\{-n\lambda(\mu-\bar{x})^2/2\}\cdot\lambda^{\alpha_n'-1}\cdot$$
$$Exp(-(\beta_n'/2)\lambda) \quad (2.37)$$

This is shown to gamma normally distribution $NG(\mu,\lambda|\mu_n',\lambda_n',\alpha_n',\beta_n')$, with $\mu_n',\lambda_n',\alpha_n',\beta_n'$ as a equality in (2.35) and (2.36)

2.5.4 Type of Statistically Enough

Definitions 2.5.4.1*Statistik T(x) statistical is called enough statistics to θmeanwhile the condition distribution X sample is given a prize to T(x) is not depend toθ.*

Definition 2.5.4.1as example is $y' = (y_1,\dots,y_n)$is a vector observation distribution depends on k parametric$\theta' = (\theta_1,\dots,\theta_k)$. Where$t' = (t_1,\dots,t_q)$is a function of q from y. than t component is called enough statistic if $\theta$for likelihood function$l(\theta|y)$and could to use in symbolize is:

$$l(\theta|y) \propto g(\theta|t), \quad (2.38)$$

And the distance result of θ, if its depend on observation, also could to called as a function of t.

Lemma 2.5.4.1. As an example t is enough to θ, which has mixing distributions $p(t|\theta)$. Than

$$l(\theta|y) \propto l_1(\theta|t) \text{ dimana}$$
$$l_1(\theta|t) \propto p(t|\theta) \quad (2.39)$$

2.6 Polya Tree

Definition: The randomize probability of *P*to*R*has a distribution to Polya Tree which has parametric is $(\Pi,\mathcal{A})$there are randomize variable *Y={Y₀, Y₀₀, Y₁₀, . . .}*if any required:

1) Randomize variables to Y independence;
2) For every$\epsilon \in E^*, Y_{\epsilon 0} \sim Beta(\alpha_{\epsilon 0},\alpha_{\epsilon 1})$;
3) For every$m = 1,2,\dots$and to every$\epsilon \in E^m$, than:

$$P(B_\epsilon) = \left[\prod_{\substack{j=1\\ \{\epsilon_j=0\}}}^{m} Y_{\epsilon_1,\dots\epsilon_j 0}\right]\left[\prod_{\substack{j=1\\ \{\epsilon_j=1\}}}^{m} 1-Y_{\epsilon_1\dots\epsilon_{j-1}0}\right] \quad (2.40)$$

Symbolizing used is

$$P \sim PT(\Pi,\mathcal{A}). \quad (2.41)$$

Randomize variable $W_1,\dots,W_n$called measurement n sample from P if:

$$P[W_i \in B_\epsilon] = EP[W_i \in B_\epsilon|P]$$
$$= E[P(B_\epsilon)]$$
$$= E[P(B_{\epsilon 1})P(B_{\epsilon 1\epsilon 2}|B_{\epsilon 1})\dots P(B_\epsilon|B_{\epsilon 1\dots\epsilon m-1})]$$
$$= \frac{\alpha_{\epsilon 1}}{\alpha_0+\alpha_1}\cdots\frac{\alpha_{\epsilon 1\dots\epsilon m-1}}{\alpha_{\epsilon 1\dots\epsilon m-1}0+\alpha_{\epsilon 1\dots\epsilon m-1}1}. \quad (2.42)$$

2.7 Markov Chain Monte Carlo (MCMC)

Markov Chain could to use to every measurement and optimizing needed. The algorithm MCMCis well known is Gibbs Sampler and the algorithmof Metropolis-Hastings (M-H).

2.7.1 Markov Chain

a. Markov Chain time discrete is a stochastic process with required discrete condition:
$P\{X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1},\dots,X_0 = i_0\} =$
$P\{X_{n+1} = j|X_n = i\}$ (2.43)
Even if
$$P_{ij} = P\{X_{n+1} = j|X_n = i\} \quad (2.44)$$
b. Markov Chain time continues is a stochastic process which has conditions of distribution from future condition with$X(t+s)$, Present$X(s)$, and past

time$X(u)$,$0 \leq u < s$,depending on present and the past time. So, They are:

$$P\{X(t+s)=j|X(s)=i, \quad X(u)=x(u), \quad 0 \leq u < s\}$$
$$= P\{X(t+s)=j|X(s)=i\} \quad (2.45)$$

Wheres, $t \geq 0$

Process of $\{X(t),\ t \geq 0\}$has a transition probability of stationers or homogeny if

$$P[X(t+s)=j|X(s)=i] = P_{ij}(t) \quad (2.46)$$

2.7.2 The Simulation Methodof Monte Carlo

Monte Carlo simulation could to use as method of statistic simulation, where statistic simulation where it's generally definition use isa statistically simulation but to all simulation method use combinations of randomize occupation number. Monte Carlo simulation is a method to evaluate one of model involve randomize occupation number as an input.

Simulasi Monte Carlo simulation could to use as method to analyze indefinitely propagation, its purpose given a randomize variation or an error influence sensitivity, from a modeling system. Monte Carlo simulation is include as a sampling method because a randomize input built up from one of probability distribution to sampling process to the real population.

2.7.3 Gibbs Sampling

Gibbssampling is an algoritmawhich basically in a series from to the fully solid condition $p(\theta_i|\theta_{\neq i}, y)$, it means posterior dense to element to-ifrom $\theta \doteq (\theta_1, \dots, \theta_d)'$, which is given by all other elements, where the element θcould similar like scalar or sub-vector.

Vector counting in a view dense as a vector of 1,2,...,s and identification j vector with j condition to the Markov Chain of the transition probability. If $i$and$j$are differencing more than in one component, Where $p_{ij} = 0$. If they are difference in maximally to one's component, for example could be concrete and different from the first component $(i \neq j)$. Use $i$vector is $(y_1, y_2, \dots, y_k)$and j vector as$(y_1^*, y_2, \dots, y_k)$. So,

$$p_{ij} = prob(Y_1 = y_1^*|Y_2 = y_{2,}Y_3 = y_3, \dots, Y_k = y_k)$$
$$= \frac{Prob\ (Y_1=y_1^*, Y_2=y_2, Y_3=y_3, \dots, Y_k=y_k)}{Prob\ (Y_2=y_2, Y_3=y_3, \dots, Y_k=y_k)} \quad (2.47)$$

The probabilities of numerator and enumerator are counts using$P_Y(y)$, we had claim that the Markov Chain is reduced an aperiodic, and then it has stationer distribution is$P_y(y)$.

3. Explanations

3.1 Annexation of unlimited Polya Tree diagram

Use the simple relative definition from Polya tree diagrams,it could be explain the generalization process in a relation of parametric distribution use to relations into N(μ, σ^2). The other relations parametric generalize by the same way or method.

The generalization process throughout from many of phases, taking J for example. In every phase it would be introduction a new parametric to generalization from the phase before it. In the first phase, lines dividing in a riil occupation number, that is endorser to the normalize distribution, than it would be divide in two intervals in μ median. And then we had observation the condition change from the top or bottom μ probability, but we have defensing of normally density form into bottom μ and the top of μ.

New parametric in the first phase is θ_{11}it is not a large than probability of μ, and θ_{12}, the probability existing in the top of μ. In formally, is correctly of X_1has a first of phase distribution, than

$$\theta_{11} \equiv P[X_1 \leq \mu], \quad (3.1)$$

And

$$\theta_{12} \equiv P[X_1 > \mu] = 1 - \theta_{11}. \quad (3.2)$$

Because of normally defending of form into both of component, if a ≤ μ and Y~ N(μ, σ^2), in formally condition is symbolize rise would be:

$$P[X_1 \leq a|X_1 \leq \mu] = \frac{P[Y \leq a]}{0.5}, \quad (3.3)$$

Where

$$P[Y \leq a] = P\left(\frac{y-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right)$$
$$= P\left(Z < \left(\frac{a-\mu}{\sigma}\right)\right)$$
$$= \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Even if

$$P[X_1 \leq a|X_1 \leq \mu] = \frac{\Phi\left(\frac{a-\mu}{\sigma}\right)}{1/2}$$
$$= 2\Phi[(a-\mu)/\sigma] \quad (3.4)$$

Where$\Phi(\cdot)$ where the function is a condition of density (cdf) from a normally standard. But in formally, if b > μ,

$$P[X_1 > b|X_1 > \mu] = \frac{P[Y > b]}{1/2}$$
$$= 2P[Y > b], \quad (3.5)$$

Within

$$P[Y > b] = P\left(\frac{y+\mu}{\sigma} > \frac{b-\mu}{\sigma}\right)$$
$$= P\left((1-Z)\left(\frac{b-\mu}{\sigma}\right)\right) = 1 - \Phi[(b-\mu)/\sigma]$$

Than

$$P[X_1 > b|X_1 > \mu] = 2P[Y > b]$$
$$= 2\{1 - \Phi b{-}\mu/\sigma \quad (3.6)$$

In the other condition we would symbolize are:

$$P[X_1 \leq a] = P[Y \leq a]2\theta_{11} \quad (3.7)$$
$$P[X_1 > b] = P[Y > b]2\theta_{12}. \quad (3.8)$$

Within $I_A(x)$ the indicator function from A, the density of the first density distribution

$$f(x_1|\mu, \sigma^2, \theta_{11}, \theta_{12}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_1-\mu)^2.2\sigma^2} \text{x } 2^1\left[\theta_{11} I_{(-\infty,\mu]}(x_1) + \theta_{12} I_{(\mu,\infty]}(x_1)\right]. \quad (3.9)$$

A given to X_2has the phase of distribution, we would explain of new parametric θ_{21}, θ_{22}, θ_{23}, θ_{24}, as a definition of relatively to useful component in the first phase:

$$\theta_{21} = P[X_2 \leq q_1|X_2 \leq \mu] \quad (3.10)$$
$$\theta_{22} = P[q_1 < X_2 \leq \mu|X_2 \leq \mu] \quad (3.11)$$
$$\theta_{23} = P[\mu < X_2 \leq q_3|X_2 > \mu] \quad (3.12)$$
$$\theta_{24} = P[q_3 < X_2|X_2 > \mu]. \quad (3.13)$$

Should to note that $\theta_{21} = 1 - \theta_{22}$and $\theta_{23} = 1 - \theta_{24}$. Without any condition, in the fourth component has:

$$P[X_2 \le q_1] = \theta_{11}\theta_{21} \quad (3.14)$$

$$P[q_1 < X_2 \le \mu] = \theta_{11}\theta_{22} \quad (3.15)$$

$$P[\mu < X_2 \le q_3] = \theta_{12}\theta_{23} \quad (3.16)$$

$$P[q_3 < X_2] = \theta_{12}\theta_{24}. \quad (3.17)$$

In every component, we had to use the original normally density is,example, if $\mu < a < b \le q_3$and Y~ N(μ, σ^2),

$$P[a < X_2 \le b]$$
$$= P[a < X_2 \le b|\mu < X_2 \le q_3]P[\mu < X_2 \le q_3]$$
$$= P[a < Y \le b|\mu < Y \le q_3]P[\mu < X_2 \le q_3]$$
$$= \frac{P[a < Y \le b]}{P[\mu < Y \le q_3]}P[\mu < X_2 \le q_3]$$
$$= P[a < Y \le b]\frac{\theta_{12}\theta_{23}}{0.25}. \quad (3.18)$$

In generally, the density of distribution in the second phases is:

$$f(x_2|\mu,\sigma^2,\theta_{11},\theta_{12},\theta_{21},\theta_{22},\theta_{23},\theta_{24}) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x_2-\mu)^2/2\sigma^2}2^2\left[\theta_{11}\theta_{21}I_{(-\infty,q_1]}(x_2) + \theta_{11}\theta_{22}I_{(q_1,\mu]}(x_2) + \theta_{12}\theta_{23}I_{(\mu,q_3]}(x_2) + \theta_{12}\theta_{24}I_{(q_3,\infty)}(x_2)\right] \quad (3.19)$$

When the Bayes analyze use with sampling distribution, we need a mixing prior distribution to parametric θ_{js}. $\theta_{js}$and then it is not difficult to interpreted, the function of prior information would be rise to them.In a extremes case selecting to $\theta_{11} = \theta_{12} = 0{,}5$within prior change 1. This case would be given a median to prior change to μ for J condition. Meanwhile, many of parametric to choose a distribution which is shown the prior information to the all of θ_{js}, even the prior information is entire. Especially,prior information would limiting by parametric from some of the first j phase.

Has a correlating, the prior distribution and posterior focus to high probability around of $\theta_{js} = 0{,}5$to all of js, this behavior is a normally distribution. This holding up when c more than to $\alpha_{js} = cp(j)$. Within $\rho(j)$ is increase, the j value in application of prior probability into θ_{js}nearest 0.5. This case an adjective which has explains first, that numerous of parametric is correlating to independent correlation.

In the other side, when small c, that its distribution more to "nonparametrc". It's given to A_J as a component into level of divide to-J. When small c, an observation into $A_J$has influence to all beta prior distribution of θ_{js} and it has correlate to A_J posterior, and it would be a high probability to $A_J$into posterior distribution. Because of A_J is a best component of divided as given to, and then main behavioral effect to discrete estimation into posterior.

The generalization sampling in the phase ofJ, or called as G, depend on θ_{js}. Prior with G to θ_{js}, a given to unlimited of diagram Polya Tree. Where the symbolize write on should to:

$$G \sim PT_j\left(c,\rho,N(\mu,\sigma^2)\right). \quad (3.20)$$

A prior to (μ, σ) that application of median μ, quartile, octal, and everything are connected,such as randomize. This case has influence to the smoothing process.

Result of sampling randomize density are built up component $\{\theta_{js}\}$ basically from these prior, but an average of prior of (μ, σ) or called as mixing diagram to Polya Tree, and its symbolize would be:

$$G \sim \int PT_j\left(c,\rho,N(\mu,\sigma^2)\right)p(d\mu,d\sigma^2). \quad (3.21)$$

Hanson (2006) shows that special prior, density of the smooth randomize MPT. The other version of Polya tree diagram, that the mixing Polya Tree diagram is unnecessary to this case; see Barron et al. (1999), and Berger also Guglielmi (2001).

3.1 Posterior Measurements

Bayes method is importing to use to non-parametric with compare to parametric analyze, these caused of uncertain ability to sampling distribution input. But meanwhile, the flexibility is built up a measurement of complexity analyze. A lot of developing of Non-Parametric Bayes modeling and has to used by simulation within measurement method, especially MCMC metode. Introduction MCMC method it is beginning from Escobar (1994) research to the mixing of Dirichlet process. In this part explained to measurements aspect from posterior sampling. This model is given to method below:

$$X_1,\ldots,X_n|G \overset{iid}{\sim} G,$$
$$G|c,\mu,\sigma \sim PT_j\left(c,\rho,N(\mu,\sigma^2)\right), \quad (3.22)$$

And

$$(\mu,\sigma^2) \sim p(\mu)p(\sigma^2).. \quad (3.23)$$

A given to c but also required to use as randomize:

4. Simulation

Polya Tree diagram and its mixed to use in data concentrate analyze. And Lavine (1994) that unlimited diagrams of Polya tree for an error model into regression control and it will show an analyze model of binomial hierarchy which has explain by Berry dan Christensen (1979). And the limit diagram of Polya Tree has success to applications into smooth modeling and GLMM (Walker and Mallick 1997) and even the distribution of error model when the break time is fast. (Walker and Mallick 1999; Mallick and Walker 2003). Hanson (2006) give to a measurement strategic to result of inferential to the mixing of limit modeling diagrams of Polya Tree, suitable to algorithm to show of modeling binomial regression with its function to non-parametric, randomize intercept modeling and various of survival model. It is similarly to the success applications purpose density estimate to the correct data (Yang, Hanson, and Christensen 2008).

It method used for to analyze an ordinal data than make it as a real data to the logistic model such as continue ratio or cumulative logic. Especially, a given data to $Y_{ij} = 1$ if i individual has an error time to j, with $Y_{ij} = 0$ if does not difficultly or an easy way as a given to a model,

$$logit\{P(Y_{ij} = 1|\beta,\gamma_i)\} = \gamma_i + \beta_1 Trt_i + \beta_2 Time_{ij} + \beta_3 Trt_i \; x \; Time_{ij} \quad (4.1)$$

Where i = 1, …, n, j = 1, …, N_i. Meanwhile γ_i is an influence randomize to each of subject, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ is a regression parametric which has correlate to Trt, this indicator is numerous binner of behavioral, times period, and interaction of time Trt x. In specially, y_i would be assumption as independence N (μ, σ^2). But it could to change into a prior normalize of MPT,

$$\gamma_1, \ldots, \gamma_n | G \overset{iid}{\sim} G,$$
$$G|\mu, \sigma^2 \sim PT_4\left(c, j^2, N(\mu, \sigma^2)\right). \qquad (4.2)$$

Normalize assumption to this methodology has influence to randomly and a false implication to the specific of normally model. Difference Prior indicator to θ_{js} given to three values of c, they are $c = 0{,}1$, $c = 1$, dan $c = 10$, level of developing shows into normally to the randomize influence. The first analyze assumption to the normalizing component are shows inconsistently among marginally and especially to subject influence and make an analyze validation suitable to normalize assumption.

Bayes analyzing needed a good prior to regression parametric $\boldsymbol{\beta}$ = (β_1, β_2, β_3)′ and also to parametric μ and σ^2 from a normal relationship to generalization by MPT. It used to normally of independen conjunctions – gamma prior invers yang menampilkan informasi prior yang lemah.

A simulation working using a software and R program, where idnr data, trt, dan time digeneret, and given to a model of equity used to (4.1) dan (4.2). the assumption parametric that β_0 = trt, β_1 = time, dan β_3 = trt x time with an other parametric is μ dan σ^2. Where given to the wellness models, and it is correction values (Deviance Information Criterion) DIC and the smaller LPML (Log Pseudo Marginal Likelihood).

The each of modeless counting used by c value. The simulation result to the each model is given by explain in table below:

Table 4.1: The simulation posterior result to the MPT model size $c = 0.1$

				95%HPD	
Parameter	Mean	Median	Std. Dev	Lower	Upper
(Intercept)	-1.8328	-1.8228	0.4552	-2.6822	-0.9701
Trt	0.2922	0.2412	0.4913	-0.6572	1.2069
Time	-0.3544	-0.3524	0.0411	-0.4291	-0.2756
Trt*time	-0.1268	-0.1253	0.0637	-0.2404	0.0017
μ	-0.7987	-0.8020	0.0589	-0.9292	0.6969
σ^2	8.4236	8.5307	1.0476	6.2112	10.1113

The simulation result to the posterior model is MPT model within $c = 0{,}1$ has resulting (Deviance Information Criterion) DIC = 930.4 and (Log Pseudo Marginal Likelihood) LPML =- 477.6

Table 4.2: The simulation posterior result to the MPT model size $c = 1$

				95%HPD	
Parameter	Mean	Median	Std. Dev	Lower	Upper
Intercept	-0.462823	-0.490515	0.530045	-1.382388	0.670885
Trt	0.136144	0.128177	0.717310	-1.250292	1.472916
Time	-0.177666	-0.166746	0.096298	-0.368031	0.006302
Trtxtime	0.029648	0.030614	0.121340	-0.191784	0.276181
μ	-0.52133	-0.55759	0.68342	-1.89266	0.79928
σ^2	0.68683	0.42639	0.85987	0.06532	2.17872

The simulation posterior result to the MPT model size c has explain (Deviance Information Criterion) DIC = 924.2 dan (Log Pseudo Marginal Likelihood) LPML = -468.0.

Table 4.3: The simulation posterior result to the MPT model size $c = 10$

				95%HPD	
Parameter	Mean	Median	Std. Dev	Lower	Upper
Intercept	-1.7524	-1.7491	0.4429	-2.5020	-0.8428
Trt	-0. 1527	-0. 1743	0.5711	-1. 2850	0. 9228
Time	-0. 3931	-0.3924	0. 0448	-0. 4898	-0.3142
Trtxtime	-0.1349	-0.1357	0.0705	-0.2730	-0.0076
μ	-1.9454	-1.9418	0.6870	-3. 3357	-0. 7314
σ^2	16. 8547	16. 2695	3. 9546	10. 4916	24. 8731

The simulation posterior result to the MPT model size $c = 10$ has result (Deviance Information Criterion) DIC = 955.3 and (Log Pseudo Marginal Likelihood) LPML = -481.3.

Table 4.4: posterior method, modeling of criteria compare, and interval 95% HPD to the mixing model of linear generalization parametric (GLMM)

	MPT		
Parameter	$\alpha = 10$	$\alpha = 1$	$\alpha = 0{,}1$
Intercept	-1.7524	-2.2108	-1.8328
Trt	-0. 1527	0.2310	0.2922
Time	-0. 3931	-0.3841	-0.3544
Trtxtime	-0.1349	-0.1301	-0.1268
μ	-1.9454	-1.2101	-0.7987
σ^2	16. 8547	22.7622	8.4236
DIC	955.3	924.2	930.4
LPML	-481.3	-468.0	-477.6
Posterior Interval			
Intercept	-2.5020, -0.8428	-3.3212, -1.2262	-2.6822, -0.9701
Trt	-1. 2850, 0.9229	-0.7307, 1.0223	-0.6572, 1.2069
Time	-0. 4898, 0.3142	-0.4735, -0.3076	-0.4291, -0.2756
Trt*time	-0.2730, 0.0076	-0.2634, -0.0020	-0.2404, 0.0017
μ	-3. 3357, 0.7314	-3.0559, 0.0173	-0.9292, 0.6969
σ^2	10. 4916, 24.8731	8.8015, 44.5144	6.2112, 10.1113

In the table 4.4 shows could explain that the useful model or the best model is in c and it has 1 value to DIC and its LPML and best of adapting model. This model is good to use than the model of c = 0.1 dan c = 10.

5. Acclusions

5.1 Conclusion

The finally conclusion difficult model which use by Bayes analysis. This case to depending on ability to use of Markov Chain Monte Carlo (MCMC) Method, and it could be to prediction of the posterior distribution. In the side of Bayes analysis is ability to giving highly flexible but meanwhile it does not use to its flexibility even these data are really needed entire.

The diagrams of Polya Tree is a randomize probability distribution, than source from MPT it could show any process of normally generalization parametric distribution. Generalization process into some of phases and we could to introduction of a new parametric to generalization from the first phases. These parametric are θ or β.

In this thesis explain of simulation used the three indicators they are c = 0,1 c =1 serta c = 10. And a given to the best model to these indicators, in fact c=1and the values of DIC and its LPML and it's called as a good adapting model. This model is better than the model of c = 0.1 dan c = 10.

5.2 Proposition

This is an interesting contraction to the diagram of Polya Tree to the multidimensionality space; and the mainly a given to the probability on distribution correlate of Polya Tree, they are the diagram of Polya Tree dependence in a space time, space, or the value of cross covariant. The Diagram of Polya Tree could be adapting and would to find out depending reality. It is suitable to the modeling of growth curves. The influence of randomize which has changed, or another application are needed the complexity of density evolutions into space time, space and its covariate.

References

[1] Bain, L.J., and Max Engelhardt (1992) Introduction to Probability and Mathematical Statistics, $2^{nd}\,edn$, Duxbury Press, California.
[2] DeGroot, M.H. (1970) Optimal Statistical Decision, McGraw-Hill.
[3] Rohatgi, V.K. (1976) An Introduction to Probability Theory and Mathematical Statistics, John Wiley-Sons.
[4] W.J.,Conover. (1971) Practical Nonparametric Statistics, John Wiley-Sons, New York-London-Sydney-Toronto.
[5] George E.P.Box and George C.Tiao. (1973) Bayesian Inference In Statistical Analyssis, Addison-Wesley Publishing Company.
[6] James O. Berger (1980) Statistical Decision Theory Foundations, and Methods, Springer-Verlag, New York Heidelberg Berlin.
[7] Mauldin,R.D.,Sudderth, W.D., and Williams, S.C. (1992). "Polya Trees and Random Distributions," *The Annals of Statistics*, 20, 1203-1221.
[8] Reich, B.J., and Fuentes, M. (2007), "Multivariate Semiparametric Bayesian Spatial Modeling Framework for Hurricane Surface Wind Fields," *Annals of Applied Statitics*, 1, 249-264.
[9] Dey,D., Muller, P., and Sinha, D. (1998), Partical Nonparametric and Semiparametric Bayesian Statistics, New York: Springer.
[10] Paddock, S.M.(1999), "Randomized Polya Trees: Bayesian Nonparametrics for Multivariate Data Analysis," Unpublished doctoral thesis, Institute of Statistics and Decision Sciences, Duke University.
[11] Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Line, A. (2000),"Bayesian Measures of Model Complexity and Fit, "*Journal of the Royal Statistical Society, Series B*, 64, 583-639.
[12] Paddock, S.M., Ruggeria, F., Lavine M., and West, M. (2003), "Randomized Polya Tree Models for Nonparametric Bayesian inference," *Statistica Sinica*, 13, 443-460.
[13] Walker, S.G., and Mallick, B.K. (1999), "Semiparametric Accelerated Life Time Model," *Biometrics*, 55, 477-483.
[14] Warren, J., Ewens, and Gregory, R., Grant. (2005), "Statisical Methods in Bioinformatics: An Introduction, Springer.